자료 분석 실습

1) Global temperature data
2) Tourist data

예제 1

- Global temperature data(1856.1~2005.12)

- 분석모형
  - 계절형 ARIMA 모형
  - 회귀모형에 의한 추세계절모형 + ARMA 오차 모형

- Training data: 1970.1~2003.12
- Test data: 2004.1~2005.12
  - 두 모형의 예측 정확성 측도 비교

예제 1

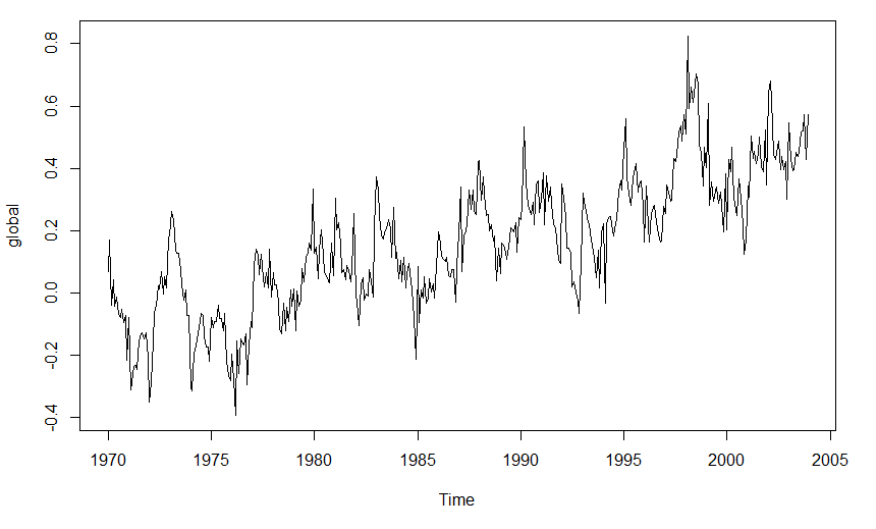
- Global temperature data(1856.1~2005.12)

- 분석모형
  - 계절형 ARIMA 모형
  - 회귀모형에 의한 추세계절모형 + ARMA 오차 모형

- Training data: 1970.1~2003.12
- Test data: 2004.1~2005.12
  - 두 모형의 예측 정확성 측도 비교

- 자료 준비

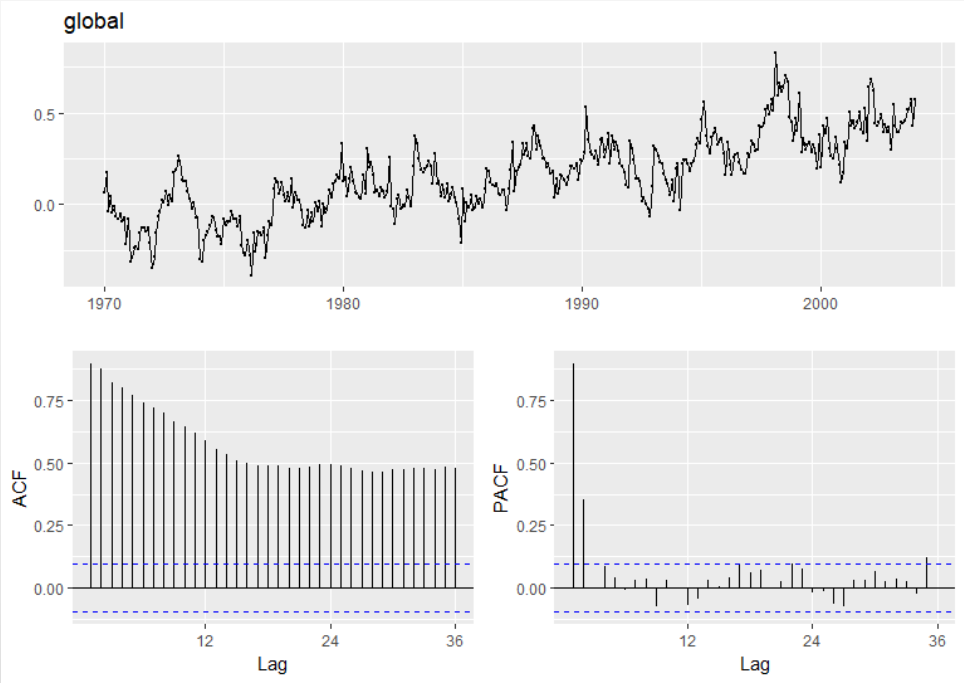
```
> Global <- scan("D:/Data/global.txt")
> Global.ts <- ts(Global, start=1856, freq=12)
> global <- window(Global.ts, start=c(1970,1), end=c(2003,12))
> new.global <- window(Global.ts, start=c(2004,1), end=c(2005,12))
> plot(global)
```



1. 계절형 ARIMA 모형

- 정상성 만족 여부 확인

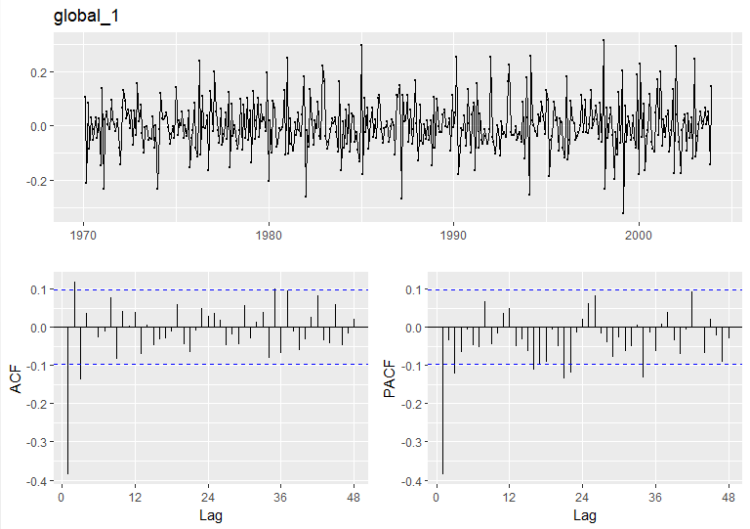
```
> library(forecast)
> ggtsdisplay(global)
```




```
> ndiffs(global)
[1] 1
> nsdiffs(global)
[1] 0
```

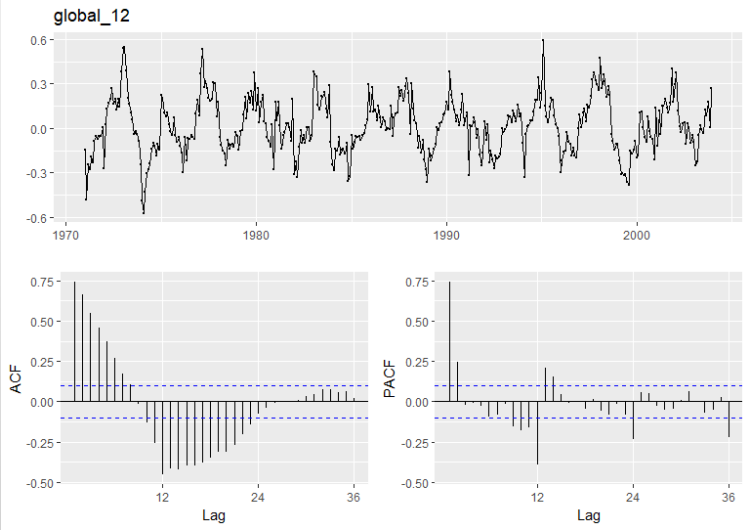
- 모형 인식

d=1



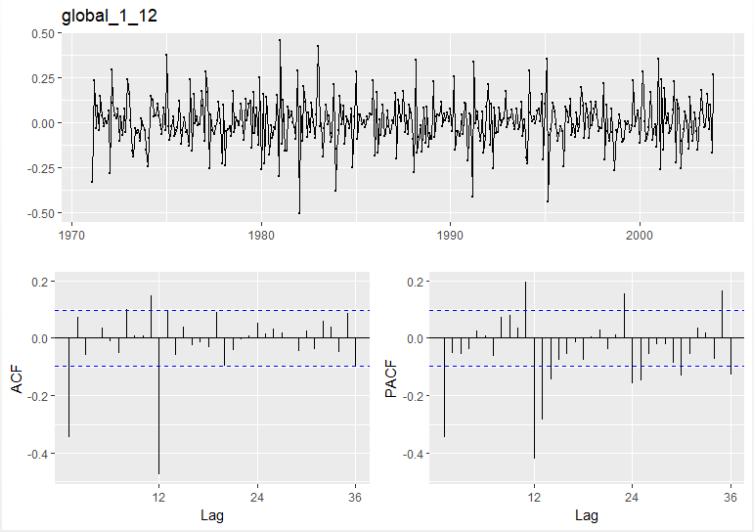
ARIMA(3,1,0)
ARIMA(0,1,3)
ARIMA(1,1,1) ARIMA(1,1,2)
ARIMA(2,1,1) ARIMA(2,1,2)

D=1



$ARIMA(2,0,0)(0,1,1)_{12}$
$ARIMA(1,0,1)(0,1,1)_{12}$ $ARIMA(2,0,1)(0,1,1)_{12}$
$ARIMA(1,0,2)(0,1,1)_{12}$ $ARIMA(2,0,2)(0,1,1)_{12}$

d=1, D=1



$ARIMA(1,1,0)(0,1,1)_{12}$
$ARIMA(0,1,1)(0,1,1)_{12}$

$ARIMA(1,1,1)(0,1,1)_{12}$
$ARIMA(2,1,1)(0,1,1)_{12}$
$ARIMA(1,1,2)(0,1,1)_{12}$
$ARIMA(2,1,2)(0,1,1)_{12}$

- d=1 모형 적합

ARIMA(3,1,0)

```
> confint(Arima(global,order=c(3,1,0),include.drift=TRUE))
               2.5 %       97.5 %
ar1   -0.502070562 -0.309184235
ar2   -0.185801325  0.022824515
ar3   -0.220353601 -0.026264016
drift -0.004202985  0.006399675
> fit1.1 <- Arima(global,order=c(3,1,0),fixed=c(NA,0,NA))
```

ARIMA(0,1,3)

```
> confint(Arima(global,order=c(0,1,3)))
           2.5 %      97.5 %
ma1 -0.50758944 -0.31328690
ma2 -0.02723333  0.17030626
ma3 -0.24132470 -0.04107896
> fit1.2 <- Arima(global,order=c(0,1,3),fixed=c(NA,0,NA))
```

```
> confint(Arima(global,order=c(1,1,1)))
              2.5 %     97.5 %
ar1  -0.5658788 0.2329961
ma1 -0.6677365 0.1613894
> confint(Arima(global,order=c(2,1,1)))
             2.5 %        97.5 %
ar1   0.4222207   0.6173191
ar2   0.2042546   0.3972461
ma1 -0.9962511 -0.9443437
> confint(Arima(global,order=c(1,1,2)))
               2.5 %       97.5 %
ar1  -1.23611788 -0.365146954
ma1 -0.08008814  0.831825148
ma2 -0.51677204  0.003929338
> confint(Arima(global,order=c(2,1,2)))
               2.5 %      97.5 %
ar1   0.02028114  0.7174530
ar2   0.15972496  0.6563412
ma1 -1.18490765 -0.4236167
ma2 -0.52107417  0.2033656
```

```
> fit1.3 <- Arima(global,order=c(2,1,1))
```

ARIMA(1,1,1): 의미 없는 모형

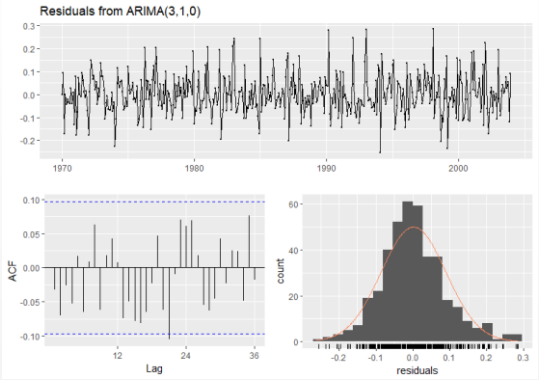
ARIMA(2,1,1): 모수 유의적

ARIMA(1,1,2): ARI(1,1)으로 축소
의미 없음

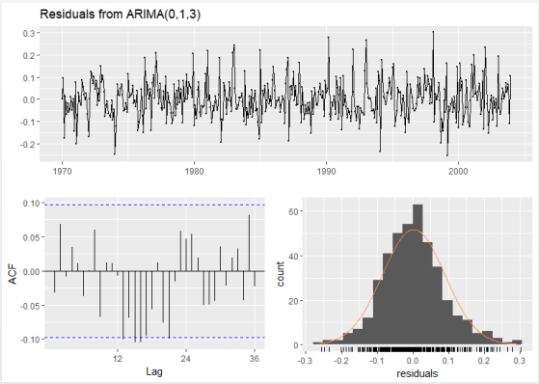
ARIMA(2,1,2): ARIMA(2,1,1)로
축소

- d=1 모형 잔차분석

```
> checkresiduals(fit1.1)

data:  Residuals from ARIMA(3,1,0)
Q* = 32.064, df = 21, p-value = 0.05768
```

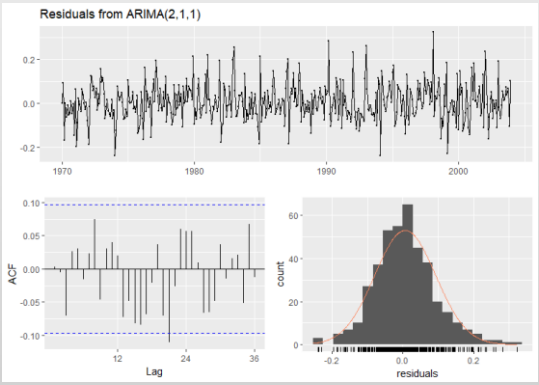



```
> checkresiduals(fit1.2)

data:  Residuals from ARIMA(0,1,3)
Q* = 37.302, df = 21, p-value = 0.01554
```



독립성 가정에 문제가 있는 모형

```
> checkresiduals(fit1.3)

data:  Residuals from ARIMA(2,1,1)
Q* = 30.228, df = 21, p-value = 0.08751
```



- 과대적합: ARIMA(3,1,0) & ARIMA(2,1,1)

```
> confint(Arima(global,order=c(3,1,1)))
          2.5 %      97.5 %
ar1 -0.4181876 0.721209072
ar2 -0.1206221 0.408492912
ar3 -0.2225624 0.021176770
ma1 -1.1403797 0.003217587
> confint(Arima(global,order=c(4,1,0)))
          2.5 %      97.5 %
ar1 -0.5103832 -0.31641445
ar2 -0.1908532  0.01799242
ar3 -0.2531151 -0.04463595
ar4 -0.1631296  0.03190238
```

추가된 모수 비유의적

- d=1에서 모형 선택

```
> c(fit1.1$aic, fit1.1$bic)
[1] -817.1217 -801.0865
> c(fit1.3$aic, fit1.3$bic)
[1] -826.5459 -810.5106
```

선택한 모형: ARIMA(2,1,1)

- D=1 모형 적합

ARIMA(2,0,0)$(0,1,1)_{12}$

```
> fit2.1 <- Arima(global,order=c(2,0,0),
                  seasonal=list(order=c(0,1,1),period=12))
> confint(fit2.1)
           2.5 %     97.5 %
ar1    0.4393444  0.6261097
ar2    0.2887728  0.4735096
sma1 -0.9603106 -0.8394875
```

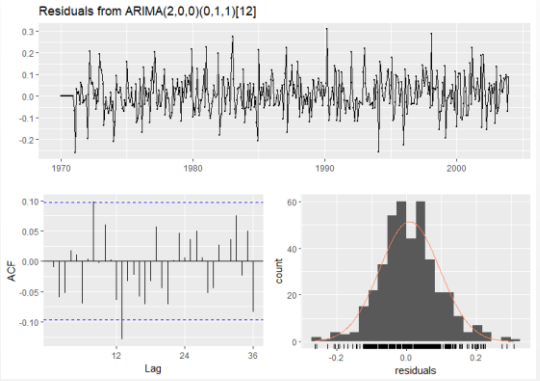
ARIMA(1,0,1)$(0,1,1)_{12}$ 선택

```
> confint(Arima(global,order=c(1,0,1),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(2,0,1),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(1,0,2),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(2,0,2),
              seasonal=list(order=c(0,1,1),period=12)))
```

```
> fit2.2 <- Arima(global,order=c(1,0,1),
                  seasonal=list(order=c(0,1,1),period=12))
```

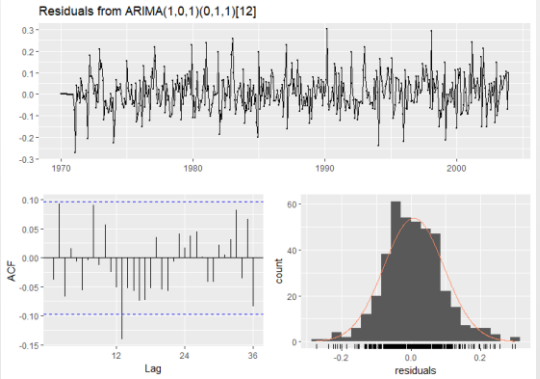
- D=1 모형 잔차 분석

```
> checkresiduals(fit2.1)

data:  Residuals from ARIMA(2,0,0)(0,1,1)[12]
Q* = 29.41, df = 21, p-value = 0.1045
```




```
> checkresiduals(fit2.2)

data:  Residuals from ARIMA(1,0,1)(0,1,1)[12]
Q* = 34.865, df = 21, p-value = 0.02921
```

독립성 가정에 문제가 있는 모형



- 과대 적합: ARIMA(2,0,0)(0,1,1)$_{12}$

```
> confint(Arima(global,order=c(3,0,0),
                seasonal=list(order=c(0,1,1),period=12)))
           2.5 %     97.5 %
ar1   0.42720979  0.6263869
ar2   0.26567593  0.4782252
ar3  -0.08238452  0.1173379
sma1 -0.96153542 -0.8403890
> confint(Arima(global,order=c(2,0,1),
                seasonal=list(order=c(0,1,1),period=12)))
           2.5 %     97.5 %
ar1   0.29404351  0.8923688
ar2   0.07052965  0.5896211
ma1  -0.38893456  0.2504489
sma1 -0.96252934 -0.8407717
```

추가된 모수 비유의적

D=1에서 선택한 모형: ARIMA(2,0,0)(0,1,1)$_{12}$

- d=1, D=1 모형 적합

```
> fit3.1 <- Arima(global,order=c(0,1,1),
                  seasonal=list(order=c(0,1,1),period=12))
> confint(fit3.1)
          2.5 %     97.5 %
ma1  -0.5148529 -0.3408150
sma1 -1.0023132 -0.8738973
```

$ARIMA(0,1,1)(0,1,1)_{12}$

```
> fit3.2 <- Arima(global,order=c(1,1,0),
                  seasonal=list(order=c(0,1,1),period=12))
> confint(fit3.2)
          2.5 %     97.5 %
ar1  -0.5103131 -0.3304573
sma1 -0.9910940 -0.8699045
```

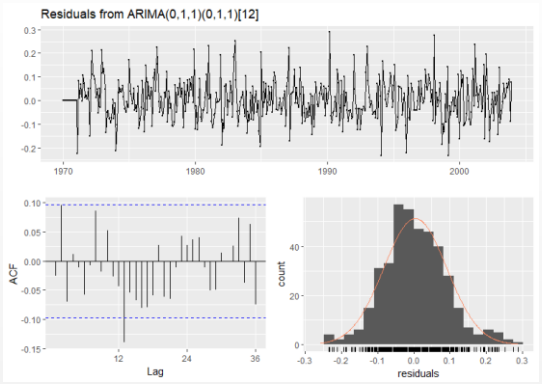
$ARIMA(1,1,0)(0,1,1)_{12}$

```
> confint(Arima(global,order=c(1,1,1),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(1,1,2),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(2,1,1),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(2,1,2),
              seasonal=list(order=c(0,1,1),period=12)))
```

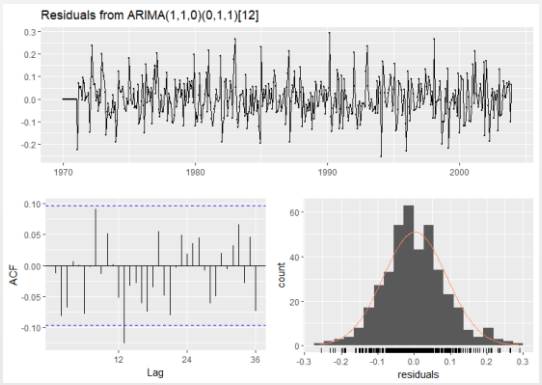
$ARIMA(2,1,1)(0,1,1)_{12}$

- 잔차 분석

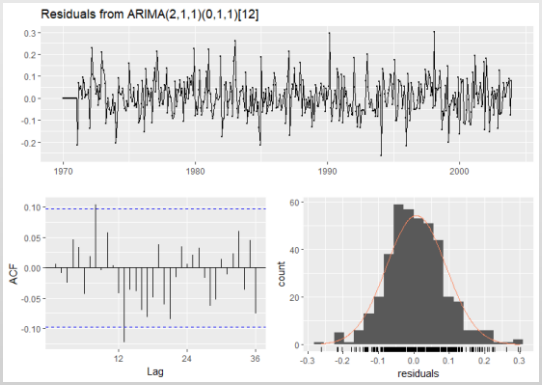
```
> checkresiduals(fit3.1)

data:  Residuals from ARIMA(0,1,1)(0,1,1)[12]
Q* = 36.564, df = 22, p-value = 0.0264
```




```
> checkresiduals(fit3.2)

data:  Residuals from ARIMA(1,1,0)(0,1,1)[12]
Q* = 31.535, df = 22, p-value = 0.08567
```




```
> checkresiduals(fit3.3)

data:  Residuals from ARIMA(2,1,1)(0,1,1)[12]
Q* = 28.778, df = 20, p-value = 0.09222
```



- 과대 적합

```
> confint(Arima(global,order=c(1,1,1),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(0,1,2),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(2,1,0),
              seasonal=list(order=c(0,1,1),period=12)))
```

추가된 모수 비유의적

```
> confint(Arima(global,order=c(2,1,2),
              seasonal=list(order=c(0,1,1),period=12)))
> confint(Arima(global,order=c(3,1,1),
              seasonal=list(order=c(0,1,1),period=12)))
```

- d=1, D=1에서 모형 선택

```
> c(fit3.1$aic, fit3.1$bic)
[1] -761.0596 -749.1229
> c(fit3.2$aic, fit3.2$bic)
[1] -762.6969 -750.7602
> c(fit3.3$aic, fit3.3$bic)
[1] -776.0254 -756.1310
```

$ARIMA(2,1,1)(0,1,1)_{12}$

- 예측 모형 비교

```
> fit1.3 <- Arima(global,order=c(2,1,1))
> fit2.1 <- Arima(global,order=c(2,0,0),
                  seasonal=list(order=c(0,1,1),period=12))
> fit3.3 <- Arima(global,order=c(2,1,1),
                  seasonal=list(order=c(0,1,1),period=12))
```

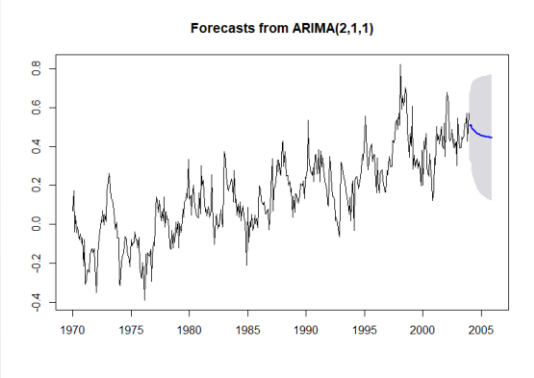
```
> AIC(fit1.3); BIC(fit1.3)
[1] -826.5459
[1] -810.5106
> AIC(fit2.1); BIC(fit2.1)
[1] -773.8951
[1] -757.9694
> AIC(fit3.3); BIC(fit3.3)
[1] -776.0254
[1] -756.131
```

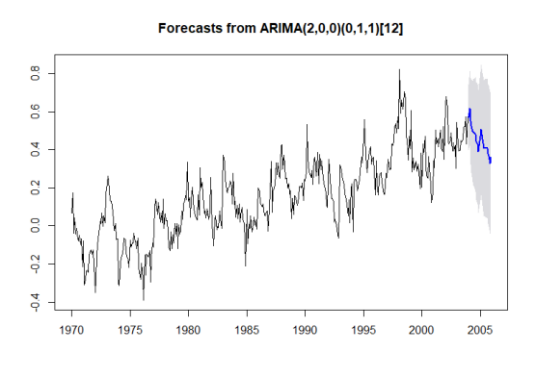
ARIMA(2,1,1): AIC, BIC 최소 모형

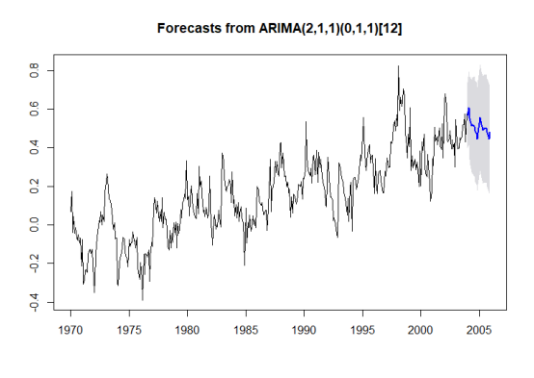
$ARIMA(2,0,0)(0,1,1)_{12}$ $ARIMA(2,1,1)(0,1,1)_{12}$: 비슷한 값의 AIC & BIC

- 예측 비교

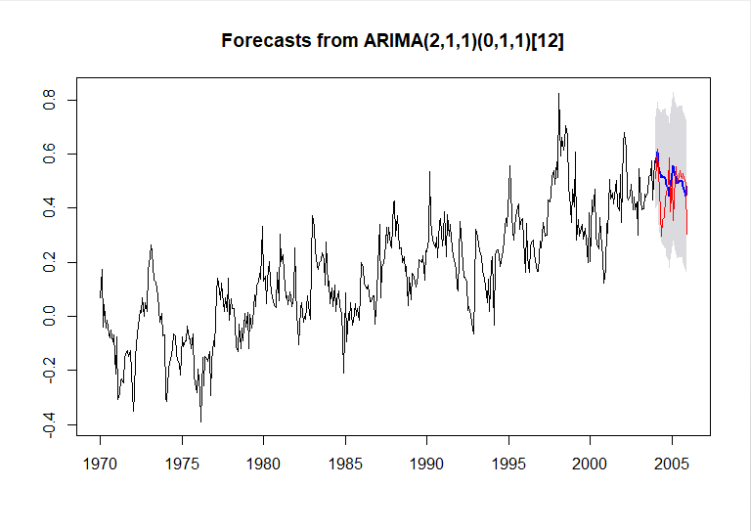
```
> fore1.3 <- forecast(fit1.3,h=2*12,level=95)
> fore2.1 <- forecast(fit2.1,h=2*12,level=95)
> fore3.3 <- forecast(fit3.3,h=2*12,level=95)
```








```
> plot(fore3.3)
> new_t <- seq(2004,by=1/12,length=24)
> lines(new_t,new.global,col="red")
```



```
> accuracy(fore1.3,new.global)
                      ME      RMSE       MAE       MPE     MAPE      MASE
Training set 0.00636976 0.0866127 0.0660101 24.33014 79.3406 0.443464
Test set     0.00465025 0.0855769 0.0731949 -2.82917 17.3306 0.491733
                    ACF1 Theil's U
Training set 0.003457635        NA
Test set     0.335253668 0.9396727

> accuracy(fore2.1,new.global)
                      ME      RMSE       MAE        MPE     MAPE      MASE
Training set 0.00901307 0.0865560 0.0659242 28.310101 78.9441 0.442887
Test set     0.01961424 0.1093193 0.0933106  0.450414 21.0868 0.626873
                    ACF1 Theil's U
Training set -0.009331956        NA
Test set      0.524759899  1.148977

> accuracy(fore3.3,new.global)
                       ME      RMSE       MAE       MPE     MAPE      MASE
Training set  0.00675948 0.0845091 0.0648402  21.2788 80.7349 0.435604
Test set     -0.03598002 0.0961873 0.0716965 -11.7109 18.2951 0.481666
                    ACF1 Theil's U
Training set 0.005439511        NA
Test set     0.305898325  1.080199
```

최종 모형: ARIMA(2,1,1)(0,1,1)$_{12}$

2. 추세계절 + ARMA 오차 회귀모형에 의한 분석

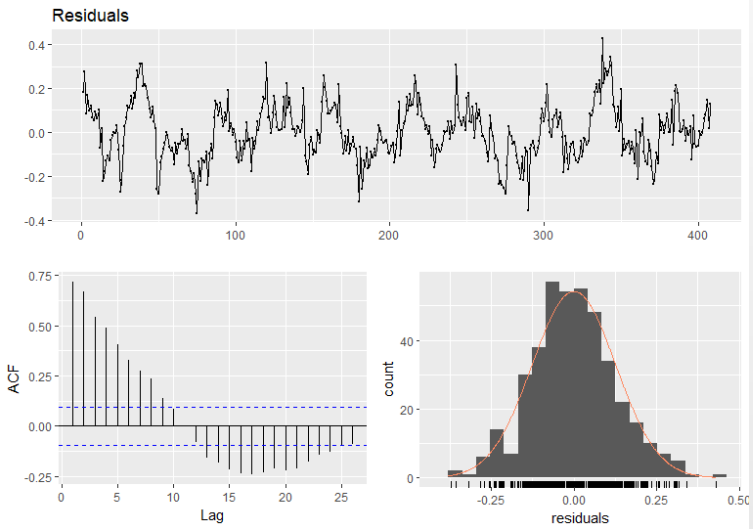
- 추세계절 회귀모형 적합

```
> Time <- time(global)
> Month <- cycle(global)
> fit1 <- lm(global~Time+factor(Month)+0)
```

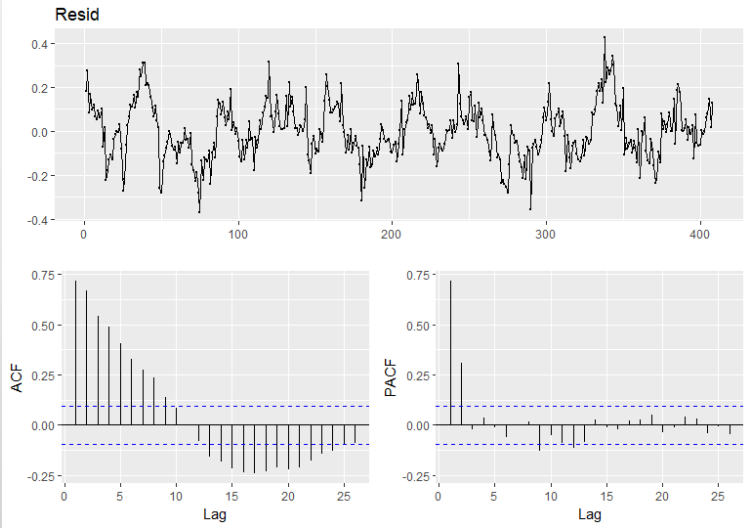
```
> summary(fit1)
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
Time              0.017759   0.000648   27.41   <2e-16 ***
factor(Month)1  -35.099567   1.287430  -27.26   <2e-16 ***
factor(Month)2  -35.088635   1.287484  -27.25   <2e-16 ***
factor(Month)3  -35.113762   1.287538  -27.27   <2e-16 ***
factor(Month)4  -35.115566   1.287592  -27.27   <2e-16 ***
factor(Month)5  -35.131310   1.287646  -27.28   <2e-16 ***
factor(Month)6  -35.124584   1.287700  -27.28   <2e-16 ***
factor(Month)7  -35.127241   1.287754  -27.28   <2e-16 ***
factor(Month)8  -35.126221   1.287808  -27.28   <2e-16 ***
factor(Month)9  -35.141436   1.287862  -27.29   <2e-16 ***
factor(Month)10 -35.157239   1.287916  -27.30   <2e-16 ***
factor(Month)11 -35.173101   1.287970  -27.31   <2e-16 ***
factor(Month)12 -35.143728   1.288024  -27.29   <2e-16 ***

Residual standard error: 0.1284 on 395 degrees of freedom
Multiple R-squared:  0.7767,  Adjusted R-squared:  0.7693
F-statistic: 105.7 on 13 and 395 DF,  p-value: < 2.2e-16
```

- 회귀모형 잔차 확인

```
> checkresiduals(fit1)

data:  Residuals
LM test = 240.46, df = 16, p-value < 2.2e-16
```




```
> Resid <- fit1$resid
> ggtsdisplay(Resid)
```



AR(2) 식별

- 오차 모형

```
> fit1_r <- Arima(Resid,order=c(2,0,0),include.mean=FALSE)
> fit1_r
Series: Resid
ARIMA(2,0,0) with zero mean

Coefficients:
         ar1     ar2
      0.4934  0.3201
s.e.  0.0469  0.0471

sigma^2 estimated as 0.006886:  log likelihood=437.15
AIC=-868.3   AICc=-868.24   BIC=-856.27
```

```
> checkresiduals(fit1_r)

data:  Residuals from ARIMA(2,0,0) with zero mean
Q* = 7.2518, df = 8, p-value = 0.5097
```

```
> confint(Arima(Resid,order=c(3,0,0),include.mean=FALSE))
> confint(Arima(Resid,order=c(2,0,1),include.mean=FALSE))
```

오차 모형: AR(2) 확정

- 추세 계절 + AR(2) 오차 회귀모형 적합

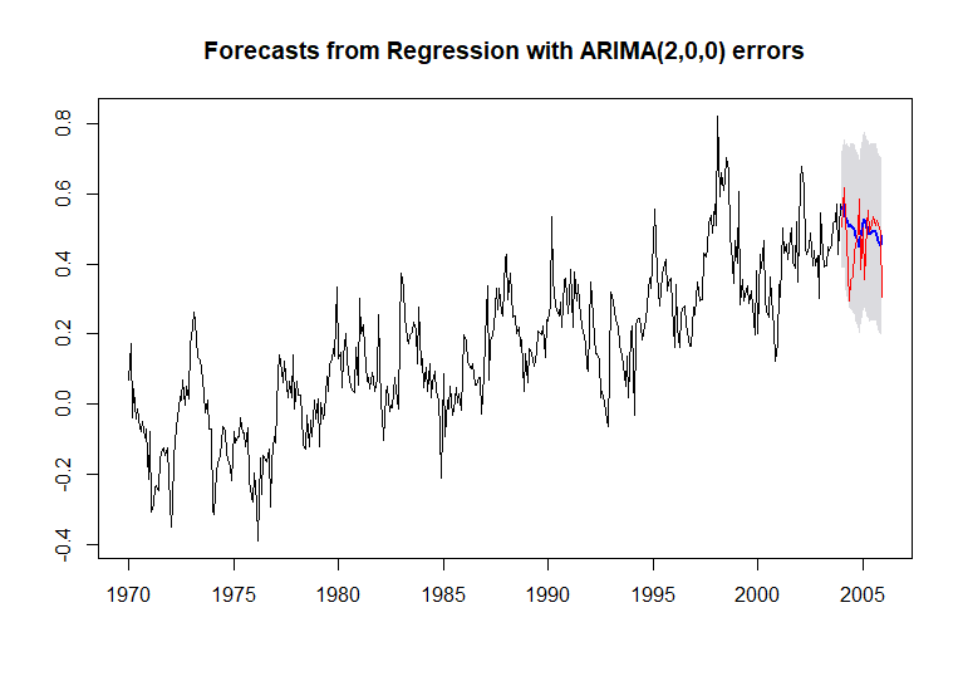
```
> fit_x <- model.matrix(fit1)
> f1 <- Arima(global, order=c(2,0,0), xreg=fit_x,
              include.mean=FALSE)
```

```
> f1
Series: global
Regression with ARIMA(2,0,0) errors

Coefficients:
         ar1     ar2    Time  factor(Month)1  factor(Month)2  factor(Month)3
      0.4932  0.3204  0.0175        -34.5101        -34.4988        -34.5242
s.e.  0.0469  0.0472  0.0021          4.2609          4.2610          4.2610
      factor(Month)4  factor(Month)5  factor(Month)6  factor(Month)7
             -34.526        -34.5417        -34.5348        -34.5373
s.e.           4.261          4.2611          4.2611          4.2611
      factor(Month)8  factor(Month)9  factor(Month)10  factor(Month)11
            -34.5359        -34.5508         -34.5660         -34.5814
s.e.           4.2612          4.2612           4.2612           4.2612
      factor(Month)12
             -34.5510
s.e.           4.2613

sigma^2 estimated as 0.007112:  log likelihood=437.21
AIC=-842.41   AICc=-841.02   BIC=-778.23
```

- 예측

```
> new.t <- time(ts(start=c(2004,1),end=c(2005,12),freq=12))
> new.x <- cbind(new.t,rbind(diag(rep(1,12)),diag(rep(1,12))))
> fore_reg <- forecast(f1,xreg=new.x,level=95)
```

```
> plot(fore_reg)
> lines(as.vector(new.t),new.global,col="red")
```



3. 예측 결과 비교

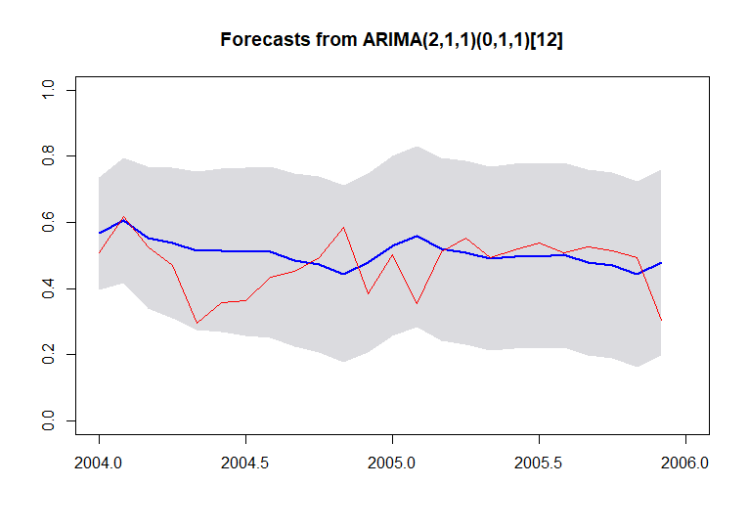
```
> fit3.3 <- Arima(global,order=c(2,1,1),
                  seasonal=list(order=c(0,1,1),period=12))
> fore_arima <- forecast(fit3.3,h=2*12,level=95)

> f1 <- Arima(global, order=c(2,0,0), xreg=fit_x,
              include.mean=FALSE)
> fore_reg <- forecast(f1,xreg=new.x,level=95)
```

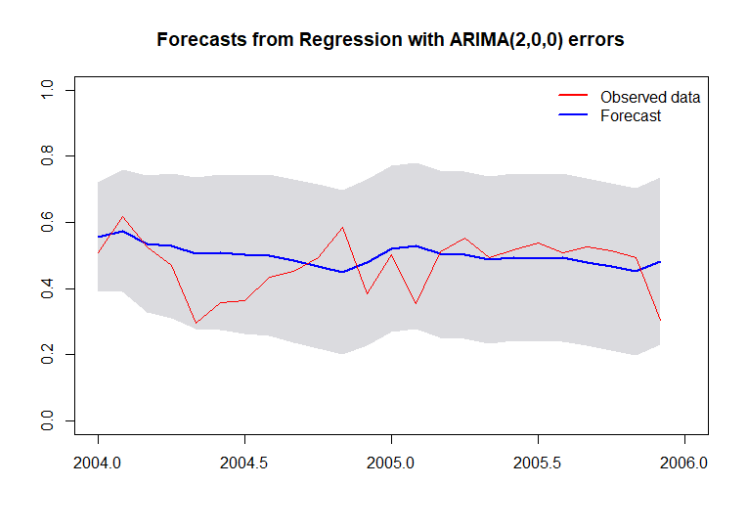
```
> accuracy(fore_reg,new.global)
                       ME      RMSE       MAE       MPE    MAPE      MASE
Training set -0.000777839 0.0827688 0.0637968  22.6275 81.5767 0.428595
Test set     -0.028508279 0.0917337 0.0694834 -10.0974 17.6225 0.466798
                    ACF1 Theil's U
Training set 0.006984407        NA
Test set     0.296332952  1.030287

> accuracy(fore_arima,new.global)
                      ME      RMSE       MAE       MPE    MAPE      MASE
Training set  0.00675948 0.0845091 0.0648402  21.2788 80.7349 0.435604
Test set     -0.03598002 0.0961873 0.0716965 -11.7109 18.2951 0.481666
                    ACF1 Theil's U
Training set 0.005439511        NA
Test set     0.305898325  1.080199
```

- 예측 그래프: 예측 기간만을 대상으로

```
> plot(fore_arima,xlim=c(2004,2006),ylim=c(0,1))
> new_t <- seq(2004,by=1/12,length=24)
> lines(new_t,new.global,col="red")
```

```
> plot(fore_reg,xlim=c(2004,2006),ylim=c(0,1))
> lines(new_t,new.global,col="red")
```






```
> legend("topright",c("Observed data","Forecast"),
        col=c("red","blue"),lwd=2,bty="n")
```